

## **Additional files**

### **Additional file 1**

#### **Structural Refinement of Modeled PA proteases**

The starting structures were refined using the Insight II equipped with DISCOVER as the energy minimization and molecular dynamics module. Structural optimization involved energy minimization (100 steps each of steepest descent and conjugate gradient methods) using cff91 force-field followed by dynamics simulations. A typical dynamics run consisted of 10000 steps of one femto-second (10 picoseconds) after 1000 steps of equilibration with a conformational sampling of 1 in 10 steps at 300K. However, dynamics simulation of 100 picoseconds was also applied to certain bigger loops for proper regularization. At the end of the dynamics simulation, the conformation with lowest potential energy was picked for the next cycle of refinement using the ANALYSIS module of Insight II. This combination of minimization and dynamics were repeated until satisfactory conformational parameters were obtained. Each loop was separately regularized applying position constraints to the rest of the atoms of the protein, which were 2 amino acids away from the desired loop by energy minimization and molecular dynamics followed by evaluation of the structural parameters. SCWRL was used to regenerate the sidechains of the modeled proteases. The final structures were energy minimized 100 steps each with steepest descent and conjugate gradient methods keeping all the atoms of the protein free.



ILE	116	1.903	4.317	5.406	2.350	-31.59	4.29	0.0000 // E=	-13.325
ASP	117	1.762	4.209	4.542	1.339	-23.44	1.09	0.0000 // E=	-10.494
ALA	118	1.411	3.018	2.048	0.027	-21.87	24.86	0.0000 // E=	9.489
PRO	119	1.742	21.173	16.628	4.890	-19.44	-35.66	0.0000 // E=	-10.673
LEU	120	1.923	3.893	2.108	0.650	-27.05	0.86	0.0000 // E=	-17.617
ASN	121	3.270	6.313	7.701	1.212	-24.63	-179.58	0.0000 // E=	-185.717
PRO	122	1.024	22.698	13.878	2.028	-15.61	11.70	0.0000 // E=	35.713
GLY	123	0.611	1.201	4.939	0.085	-19.43	28.94	0.0000 // E=	16.340
ASN	124	2.181	13.704	2.973	1.796	-39.24	-188.51	0.0000 // E=	-207.093
SER	125	1.526	1.729	2.233	0.086	-21.82	-1.81	0.0000 // E=	-18.057
GLY	126	0.814	3.499	1.309	0.376	-27.23	68.95	0.0000 // E=	47.718
GLY	127	0.244	1.638	3.344	0.386	-15.34	51.67	0.0000 // E=	41.940
PRO	128	3.139	41.645	12.576	0.600	-19.96	-32.10	0.0000 // E=	5.892
ALA	129	1.853	1.228	3.445	0.686	-35.58	-20.16	0.0000 // E=	-48.520
LEU	130	6.058	40.164	9.688	0.374	-31.09	-18.42	0.0000 // E=	6.768
VAL	131	2.693	14.937	12.119	3.692	-4.23	-9.10	0.0000 // E=	20.103
ARG	132	6.130	65.507	7.708	1.585	-1.79	-215.89	0.0000 // E=	-136.750
GLY	133	2.951	2.115	1.714	0.762	-12.28	40.49	0.0000 // E=	35.747
LYSH	134	70.535	114.694	40.086	1.966	58.86	-8.14	0.0000 // E=	278.003
VAL	135	48.193	296.793	53.665	50.548	91.59	0.11	0.0000 // E=	540.900
VAL	136	5.453	16.108	1.843	4.679	9.57	40.11	0.0000 // E=	77.754
GLY	137	1.011	3.954	4.211	0.853	-13.89	32.04	0.0000 // E=	28.179
ILE	138	3.134	5.440	8.234	2.386	-21.89	101.83	0.0000 // E=	99.139
-----									
KJ/mol		415.064	1872.979	1086.253	339.968	-2699.08	-1965.17	0.0000 // E=	-949.983

// Computations were done in vacuo with the GROMOS96 43B1 parameters set, without reaction field.  
// For more information about GROMOS96, refer to: W.F. van Gunsteren et al. (1996) in Biomolecular  
// simulation: the GROMOS96 manual and user guide. Vdf Hochschulverlag ETHZ (<http://igc.ethz.ch/gromos>).  
// When using those results, please mention that energy computations were done with the GROMOS96  
// implementation of Swiss-PdbViewer.



ILE A 120	0.930	5.300	4.068	1.546	-22.55	6.19	0.0000 // E=	-4.521
SER A 121	0.175	3.055	8.459	1.334	-22.34	-6.00	0.0000 // E=	-15.316
THR A 122	1.526	7.614	19.131	2.523	-26.36	-34.52	0.0000 // E=	-30.086
SER A 123	3.540	10.854	8.793	2.653	-7.70	-14.95	0.0000 // E=	3.192
LEU A 124	2.586	36.436	2.970	3.442	3.98	-1.69	0.0000 // E=	47.724
GLU A 125	0.697	21.172	22.412	12.800	-20.44	-1.40	0.0000 // E=	35.244
ILE A 126	1.783	7.681	21.718	1.228	-4.56	4.69	0.0000 // E=	32.531
ARG A 127	1.751	9.209	10.783	1.923	-41.34	-270.51	0.0000 // E=	-288.192
THR A 128	0.926	2.771	14.797	2.024	-14.11	-15.66	0.0000 // E=	-9.252
THR A 129	1.266	2.016	7.369	0.343	-28.57	-32.64	0.0000 // E=	-50.220
MET A 130	3.137	9.139	8.511	3.306	-47.94	-2.29	0.0000 // E=	-26.137
GLU A 131	8.868	28.028	12.655	5.823	-44.44	3.59	0.0000 // E=	14.516
VAL A 132	10.686	50.218	25.592	14.556	-19.56	4.54	0.0000 // E=	86.042
ALA A 133	1.205	3.217	2.128	2.474	-5.89	27.95	0.0000 // E=	31.085
PRO A 134	0.639	17.987	27.025	1.727	-29.46	13.83	0.0000 // E=	31.742
GLY A 135	0.327	0.473	6.345	0.527	-9.45	41.15	0.0000 // E=	39.375
ASP A 136	0.695	2.455	4.249	0.407	-37.25	-3.73	0.0000 // E=	-33.175
SER A 137	0.974	1.425	2.082	0.538	-26.52	29.27	0.0000 // E=	7.764
GLY A 138	0.884	8.413	2.335	1.787	-19.71	42.67	0.0000 // E=	36.375
SER A 139	2.304	4.013	1.045	2.805	-26.05	24.41	0.0000 // E=	6.531
PRO A 140	0.817	16.490	17.133	3.001	-42.71	-31.48	0.0000 // E=	-36.747
VAL A 141	2.126	3.870	4.229	1.062	-24.69	4.47	0.0000 // E=	-8.933
PHE A 142	2.496	12.244	5.340	2.358	-53.86	84.16	0.0000 // E=	52.738
<hr/>								
KJ/mol	253.108	1121.079	1195.992	311.554	-3750.23	-1639.93	0.0000 // E=	-2508.426

// Computations were done in vacuo with the GROMOS96 43B1 parameters set, without reaction field.  
// For more information about GROMOS96, refer to: W.F. van Gunsteren et al. (1996) in Biomolecular  
// simulation: the GROMOS96 manual and user guide. Vdf Hochschulverlag ETHZ (<http://igc.ethz.ch/gromos>).  
// When using those results, please mention that energy computations were done with the GROMOS96  
// implementation of Swiss-PdbViewer.



TYR	A 120	3.469	29.658	7.084	17.527	-52.19	-43.14	0.0000 // E=	-37.594
GLN	A 121	2.141	11.624	4.288	14.672	-33.53	-166.01	0.0000 // E=	-166.815
ALA	A 122	0.789	2.647	0.358	0.084	-21.32	-3.90	0.0000 // E=	-21.336
SER	A 123	2.424	3.790	4.872	0.426	-14.95	-18.57	0.0000 // E=	-22.008
ALA	A 124	1.668	3.981	3.294	0.026	-13.36	-8.08	0.0000 // E=	-12.474
ALA	A 125	0.978	2.677	2.939	0.889	-10.48	0.09	0.0000 // E=	-2.901
ALA	A 126	1.214	1.335	4.065	0.445	-16.19	-4.63	0.0000 // E=	-13.761
SER	A 127	1.841	3.573	2.801	0.977	-19.19	17.61	0.0000 // E=	7.616
GLY	A 128	1.096	1.440	0.740	0.841	-8.52	74.75	0.0000 // E=	70.343
GLY	A 129	0.810	7.580	2.943	0.356	-10.00	32.57	0.0000 // E=	34.256
SER	A 130	2.139	3.251	8.597	1.057	-28.43	-22.74	0.0000 // E=	-36.128
SER	A 131	1.517	6.167	3.090	0.723	-10.33	36.54	0.0000 // E=	37.698
GLY	A 132	0.390	8.780	1.612	0.026	-18.39	40.53	0.0000 // E=	32.946
SER	A 133	3.137	7.178	15.136	0.140	-21.30	9.08	0.0000 // E=	13.379
PRO	A 134	14.174	56.754	8.696	3.911	33.67	-28.90	0.0000 // E=	88.309
VAL	A 135	2.116	4.553	9.345	0.167	-27.42	3.97	0.0000 // E=	-7.264
VAL	A 136	82.716	103.096	3.095	13.546	155.70	-18.54	0.0000 // E=	339.608
ASN	A 137	20.026	4.533	5.266	0.982	-14.04	-150.70	0.0000 // E=	-133.933
LYS	A 138	5.166	36.285	15.670	0.724	17.72	1.74	0.0000 // E=	77.308
ASP	A 139	0.374	2.938	15.779	0.334	-16.91	47.86	0.0000 // E=	50.379
GLY	A 140	0.769	0.527	0.598	0.088	-12.80	46.35	0.0000 // E=	35.531
PHE	A 141	18.218	33.459	7.372	14.247	19.91	2.31	0.0000 // E=	95.516
ALA	A 142	0.766	1.132	1.965	0.594	-19.92	-8.87	0.0000 // E=	-24.337
VAL	A 143	26.803	53.610	4.298	2.344	87.01	-2.81	0.0000 // E=	171.256
ALA	A 144	3.913	5.589	0.614	1.041	-9.98	2.23	0.0000 // E=	3.411
LEU	A 145	1.527	9.781	9.456	0.849	-46.43	5.84	0.0000 // E=	-18.982
GLN	A 146	9.962	58.995	5.616	25.445	-26.53	-184.45	0.0000 // E=	-110.961
ALA	A 147	2.183	15.903	5.336	16.918	-8.54	50.57	0.0000 // E=	82.378
GLY	A 148	0.838	1.902	7.567	1.227	-5.88	91.45	0.0000 // E=	97.109
GLY	A 149	0.471	2.874	2.440	0.190	-13.27	54.00	0.0000 // E=	46.701
ARG	A 150	1.664	4.488	8.298	2.545	-20.63	-261.49	0.0000 // E=	-265.121
ALA	A 151	0.200	1.231	5.133	0.261	-18.06	-1.14	0.0000 // E=	-12.382
ASP	A 152	0.311	5.790	12.891	0.599	-12.57	44.22	0.0000 // E=	51.240
GLY	A 153	0.157	3.140	7.036	0.047	-9.67	46.16	0.0000 // E=	46.867
ALA	A 154	0.428	2.270	27.995	2.020	-10.72	8.13	0.0000 // E=	30.120
SER	A 155	1.559	12.128	27.903	1.588	1.47	49.30	0.0000 // E=	93.950
-----									
KJ/mol		592.644	2001.879	1315.600	453.844	-2700.92	-2962.41	0.0000 // E=	-1299.371

// Computations were done in vacuo with the GROMOS96 43B1 parameters set, without reaction field.  
// For more information about GROMOS96, refer to: W.F. van Gunsteren et al. (1996) in Biomolecular  
// simulation: the GROMOS96 manual and user guide. Vdf Hochschulverlag ETHZ (<http://igc.ethz.ch/gromos>).  
// When using those results, please mention that energy computations were done with the GROMOS96  
// implementation of Swiss-PdbViewer.

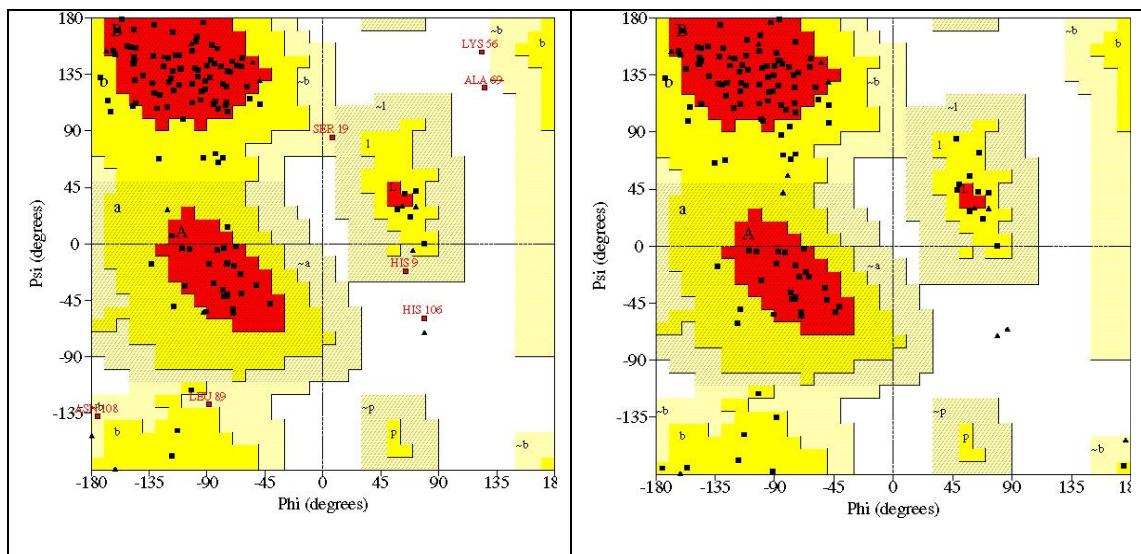


THR	A 119	1.234	4.416	4.295	1.871	-10.01	17.12	0.0000 // E=	18.925
GLY	A 120	0.797	0.761	11.439	0.217	-16.71	26.09	0.0000 // E=	22.593
VAL	A 121	1.323	4.840	4.536	2.677	-19.41	-8.70	0.0000 // E=	-14.733
THR	A 122	2.043	4.078	3.294	2.587	-13.14	-6.16	0.0000 // E=	-7.291
ILE	A 123	3.609	6.516	4.042	1.982	-9.49	46.00	0.0000 // E=	52.653
GLY	A 124	0.901	2.469	0.333	0.522	-20.36	70.67	0.0000 // E=	54.533
GLY	A 125	0.796	9.857	3.525	1.022	-17.58	72.61	0.0000 // E=	70.229
GLY	A 126	0.431	3.868	1.406	0.528	-17.47	28.56	0.0000 // E=	17.329
ILE	A 127	2.742	5.541	11.505	1.898	-33.01	-9.26	0.0000 // E=	-20.584
GLN	A 128	2.305	5.672	5.253	0.873	-32.41	-173.62	0.0000 // E=	-191.921
THR	A 129	2.090	6.129	1.741	2.006	-36.54	-21.96	0.0000 // E=	-46.530
ASP	A 130	1.190	4.546	8.472	0.677	-17.09	5.55	0.0000 // E=	3.347
ALA	A 131	0.357	1.354	3.376	0.024	-22.73	2.23	0.0000 // E=	-15.397
ALA	A 132	0.788	1.485	1.080	0.428	-15.25	-9.31	0.0000 // E=	-20.779
ILE	A 133	2.408	3.583	3.438	2.753	-35.17	-2.35	0.0000 // E=	-25.337
ASN	A 134	2.099	6.562	2.949	1.714	-29.39	-160.06	0.0000 // E=	-176.118
PRO	A 135	0.682	19.326	25.241	0.798	-8.64	14.17	0.0000 // E=	51.577
GLY	A 136	0.476	1.166	1.384	0.095	-17.17	44.77	0.0000 // E=	30.720
ASN	A 137	1.716	6.167	5.208	0.678	-39.86	-162.02	0.0000 // E=	-188.108
SER	A 138	1.954	7.014	5.441	0.512	-28.44	16.57	0.0000 // E=	3.049
GLY	A 139	1.310	8.209	0.229	1.013	-23.38	70.34	0.0000 // E=	57.721
GLY	A 140	1.899	6.554	2.283	0.077	-19.91	44.02	0.0000 // E=	34.923
PRO	A 141	1.637	20.334	18.042	0.355	-41.05	-27.13	0.0000 // E=	-27.805
LEU	A 142	2.124	10.805	4.048	0.762	-46.16	-3.76	0.0000 // E=	-32.180
LEU	A 143	1.339	12.023	5.476	0.300	-46.33	-14.11	0.0000 // E=	-41.303
ASP	A 144	1.768	4.373	10.505	4.522	-31.33	11.25	0.0000 // E=	1.080
SER	A 145	0.784	7.705	1.337	2.090	-9.87	-10.40	0.0000 // E=	-8.357
LYS	A 146	1.439	17.121	6.329	0.371	-18.29	46.20	0.0000 // E=	53.177
GLY	A 147	0.883	2.723	9.576	0.732	-19.36	48.66	0.0000 // E=	43.211
ASN	A 148	2.913	6.295	7.454	0.994	-38.09	-168.49	0.0000 // E=	-188.917
LEU	A 149	1.384	7.805	3.671	0.995	-33.57	-5.28	0.0000 // E=	-24.997
ILE	A 150	1.791	8.368	34.776	4.396	-30.37	38.11	0.0000 // E=	57.070
GLY	A 151	0.381	3.970	0.335	1.028	-16.97	40.34	0.0000 // E=	29.086
ILE	A 152	2.221	7.982	16.448	0.672	-25.37	-3.65	0.0000 // E=	-1.694
ASN	A 153	2.566	4.050	14.660	0.237	-50.72	-187.55	0.0000 // E=	-216.758
THR	A 154	3.161	5.912	0.861	0.830	-37.48	-14.62	0.0000 // E=	-41.328
ALA	A 155	0.905	1.495	0.502	0.403	-25.03	-3.92	0.0000 // E=	-25.646
ILE	A 156	1.701	3.582	2.258	1.866	-26.11	-5.59	0.0000 // E=	-22.296
PHE	A 157	0.509	4.950	6.867	0.634	-25.93	5.59	0.0000 // E=	-7.383
THR	A 158	2.289	3.277	26.916	1.812	-15.97	-18.59	0.0000 // E=	-0.269
GLN	A 159	4.351	56.755	6.627	10.858	-7.32	-160.22	0.0000 // E=	-88.955
THR	A 160	0.786	49.027	28.239	13.191	5.60	32.19	0.0000 // E=	129.034
GLY	A 161	0.649	2.128	2.794	0.061	-12.92	31.43	0.0000 // E=	24.138
THR	A 162	0.301	1.217	3.859	0.609	-16.82	-10.27	0.0000 // E=	-21.110
SER	A 163	0.315	4.459	4.160	1.106	-13.99	-4.04	0.0000 // E=	-7.992
ALA	A 164	0.518	0.690	13.541	0.728	-11.32	30.28	0.0000 // E=	34.433
GLY	A 165	0.281	1.244	3.968	0.010	-13.10	40.67	0.0000 // E=	33.077
VAL	A 166	0.458	1.459	0.349	1.226	-20.90	27.94	0.0000 // E=	10.528
GLY	A 167	0.369	0.703	2.306	0.007	-16.74	20.65	0.0000 // E=	7.299
PHE	A 168	0.570	1.709	5.223	0.513	-47.51	-5.42	0.0000 // E=	-44.921
ALA	A 169	1.207	1.938	0.483	0.575	-36.04	-10.23	0.0000 // E=	-42.068
ILE	A 170	5.109	7.620	3.105	1.928	-29.46	17.90	0.0000 // E=	6.203
PRO	A 171	2.664	19.064	22.639	2.377	-22.81	31.53	0.0000 // E=	55.465

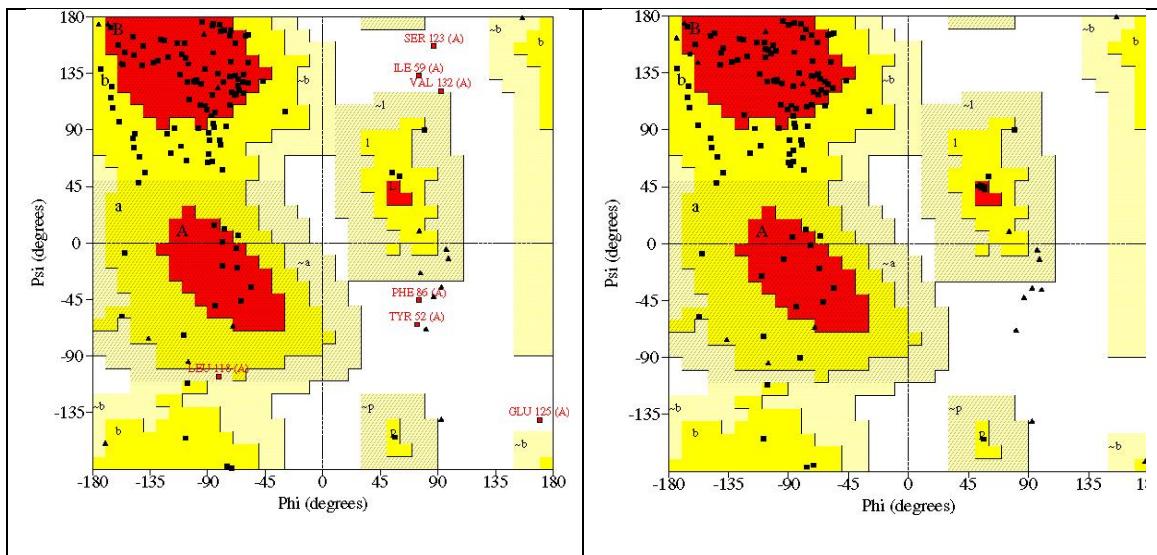
---

KJ/mol      291.422      1598.663      1432.665      358.775      -4068.57      -2896.47      0.0000 // E=      -3283.516

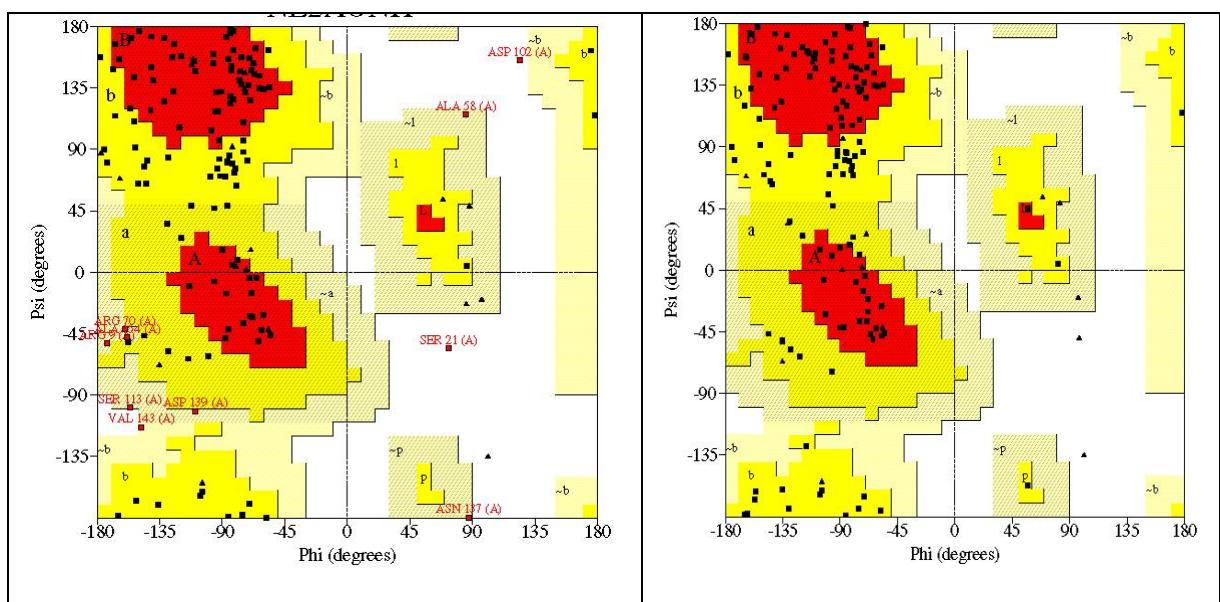
// Computations were done in vacuo with the GROMOS96 43B1 parameters set, without reaction field.  
 // For more information about GROMOS96, refer to: W.F. van Gunsteren et al. (1996) in Biomolecular  
 // simulation: the GROMOS96 manual and user guide. Vdf Hochschulverlag ETHZ (<http://igc.ethz.ch/gromos>).  
 // When using those results, please mention that energy computations were done with the GROMOS96  
 // implementation of Swiss-PdbViewer.



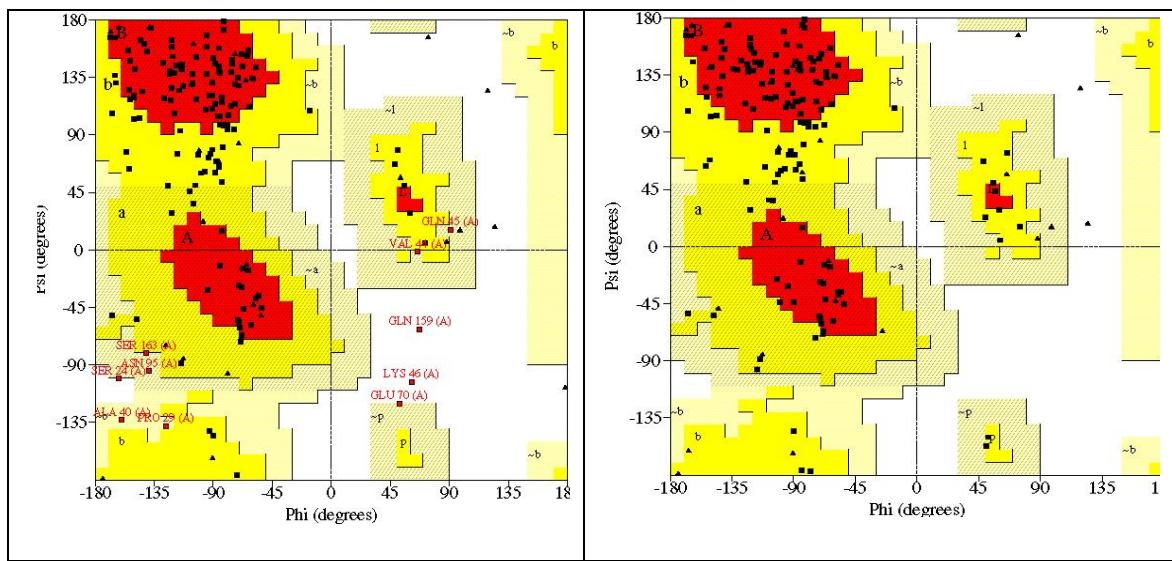
**Figure S1. Ramachandran plot of  $\phi$ - $\psi$  dihedral angles of a modeled PA serine protease structure from *Plasmodium falciparum* before and after backbone refinement.** PROCHECK was used to check the distribution of  $\phi$ - $\psi$  dihedral angles and eliminate Ramachandran outliers in the modeled protease structure (A, before; B, after refinement). Residues whose  $\phi$ - $\psi$  pairs fell outside the most favourable (red) and additional allowed (yellow) zones are annotated in red.



**Figure S2. Ramachandran plot of  $\phi$ - $\psi$  dihedral angles of a modeled PA serine protease structure from *Pyrococcus furiosus* before and after backbone refinement.** PROCHECK was used to check the distribution of  $\phi$ - $\psi$  dihedral angles and eliminate Ramachandran outliers in the modeled protease structure (A, before; B, after refinement). Residues whose  $\phi$ - $\psi$  pairs fell outside the most favourable (red) and additional allowed (yellow) zones are annotated in red.



**Figure S3. Ramachandran plot of  $\phi$ - $\psi$  dihedral angles of a modeled PA serine protease structure from *Neurospora crassa* before and after backbone refinement.** PROCHECK was used to check the distribution of  $\phi$ - $\psi$  dihedral angles and eliminate Ramachandran outliers in the modeled protease structure (A, before; B, after refinement). Residues whose  $\phi$ - $\psi$  pairs fell outside the most favourable (red) and additional allowed (yellow) zones are annotated in red.



**Figure S4. Ramachandran plot of  $\phi$ - $\psi$  dihedral angles of a modeled PA serine protease structure from *Arabidopsis thaliana* before and after backbone refinement.** PROCHECK was used to check the distribution of  $\phi$ - $\psi$  dihedral angles and eliminate Ramachandran outliers in the modeled protease structure (A, before; B, after refinement). Residues whose  $\phi$ - $\psi$  pairs fell outside the most favourable (red) and additional allowed (yellow) zones are annotated in red.

**Table S5. Predicted hydrogen bonds in modeled PA protease structures**

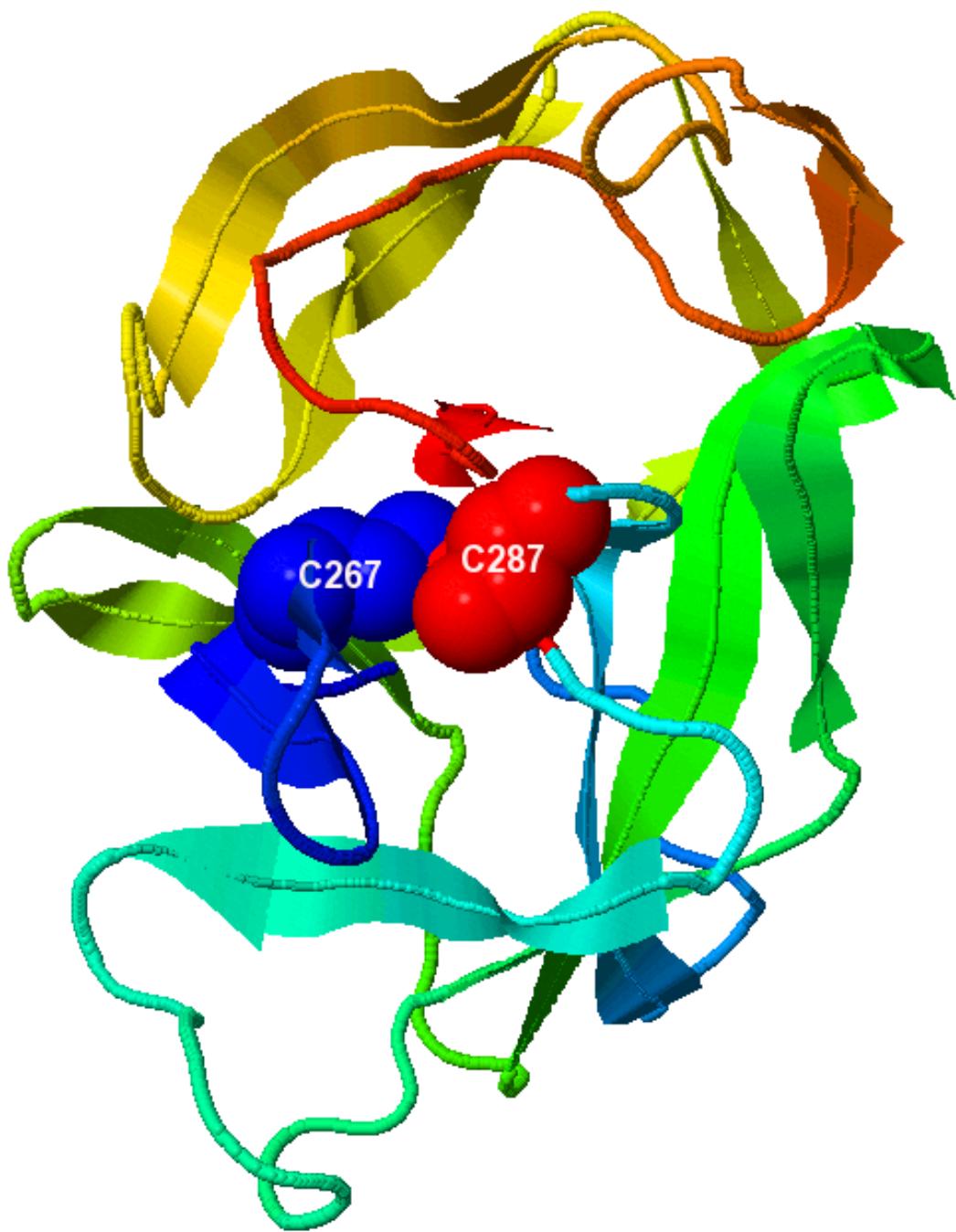
<i>B. taurus</i> ( <i>CTRB</i> ), for numbering reference	<i>P. falciparum</i> (PM0075793)	<i>P. furiosus</i> (PM0075794)	<i>N. crassa</i> (PM0075795)	<i>A. thaliana</i> (PM0075796)
Thr54	Thr325 (Ser314, 2.69 Å)	Thr283 (Ala302, 2.90 Å)	Thr117 (Gly153, 2.83 Å) (Asp189, 2.95 Å)	Thr96 (Ala144, 2.87 Å)
Ala56	Ala327 (Ser331, 2.88 Å) (Asp359, 2.99 Å)			Ala98 (Asp130, 2.84 Å)
His57	<b>His328</b>	<b>His286</b> (Asp320, 2.87 Å)	<b>His120</b> (Asp151, 2.45 Å)	<b>His99</b>
<b>Asp102</b>	<b>Asp359</b> (Ala327, 2.99 Å)	<b>Asp320</b> (His286, 2.87 Å) (Gly285, 2.76 Å) (Thr381, 2.92 Å)	<b>Asp151</b> (His120, 2.45 Å)	<b>Asp130</b> (Ala98, 2.84 Å)
Gly193	Gly436 (Ser314, 2.95 Å) (Asp433, 2.90 Å)	Gly387	Gly222	
Asp194		Asp388 (Arg360, 2.92 Å) (Thr361, 2.67 Å)		
<b>Ser195</b>	<b>Ser438</b> (Ser314, 2.79 Å) (Asp329, 2.70 Å) (Asp433, 2.91 Å)	<b>Ser389</b> (Thr268, 2.92 Å)	<b>Ser234</b> (Gln249, 2.63 Å) (Ser236, 2.81 Å)	<b>Ser208</b> (Ser84, 2.63 Å) (Asn216, 2.93 Å)
Gly196	Gly439	Gly390	Gly235 (Gly2, 2.59 Å)	Gly209
Gly197	Gly440 (Asp436, 2.97 Å)			
Pro198	Pro441	Pro392	Pro237	

DeepView/Swiss-PdbViewer was used to calculate potential hydrogen bonds between residues of modeled PA protease structures with either a hydrogen molecule present (1.20-2.76 Å, minimum angle 120°) or not present (2.19-3.00 Å, minimum angle 90°). Highly conserved amino acids (including catalytic triad residues in bold) from the multiple sequence alignment are listed with potential binding partners, including the minimum bond length (Å). Bovine chymotrypsin B (*CTRB*) is included as a standard reference for residue numbering.

**Table S6. Disulfide bonds in close proximity to catalytic histidine residue of experimental structures and modeled structures of PA serine proteases**

PA protease structure	Disulfide bond pairs	Distance in amino acids between cysteine residues
1A0L	(C59 ↔ C75)	16
1JRS	(C48 ↔ C64)	16
1DPO	(C48 ↔ C64)	16
1BIT	(C45 ↔ C61)	16
1AO5	(C50 ↔ C66)	16
1BQY	(C50 ↔ C66)	16
1SGI	(C391 ↔ C407)	16
1ABJ	(C391 ↔ C407)	16
2ANY	(C419 ↔ C435)	16
1TRY	(C50 ↔ C66)	16
1EKB	(C826 ↔ C842)	16
<b>1M9U</b>	<b>(C29 ↔ C45) 2.03 Å</b>	<b>16</b>
1EQ9	(C26 ↔ C42)	16
1ARC	(C241 ↔ C263)	22
1QY6	-	-
1SGC	(C130 ↔ C150)	20
PM0075796	-	-
<b>PM0075794</b>	<b>(C267 ↔ C287) 2.04 Å</b>	<b>20</b>
PM0075793	-	-
PM0075795	-	-

For experimental PA serine protease structures, a disulfide bond in close proximity to the catalytic histidine residue has been previously described and is annotated in UniProt (with the exception of 1M9U). For the modeled PA serine proteases, only PM0075794 (*P. furiosus*) had a cysteine residue close to the catalytic histidine. Based on homology, we predict a disulfide bond between C29 and C45 (2.03 Å) of the M9U structure (*E. fetida*) and between C267 and C287 2.04 Å of the PM0075794 modeled structure (as shown in Figure S5).



**Figure S5. Predicted disulfide bond in Modeled PA protease structure of *Pyrococcus furiosus* (PMDB ID: PM0075794).** The ribbon model shows secondary structures ( $\beta$ -sheets with arrow directed to C-terminus,  $\alpha$ -helices and turn/loops) in alternating colors and cysteine residues Cys 267 (blue) and Cys287 (red) forming a predicted disulfide bond (2.04 Å).

**Table S7. Relative comparison of PA serine protease amino acid composition based on physico-chemical properties**

ID	Small (%)	Aliphatic (%)	Aromatic (%)	Nonpolar (%)	Polar (%)	Charged (%)	Basic (%)	Acidic (%)
<b>All residues</b>								
<i>CTR</i> B	43	24	12	45	55	31	19	12
1ARC-A	67	16	9	54	46	18	9	8
1QY6-A	63	16	9	46	54	22	10	12
1SGC-A	69	18	8	58	42	13	8	5
1TRY-A	72	20	7	61	39	10	6	4
1EKB-B	54	21	12	54	46	20	8	11
1JRS-A	63	21	9	57	43	14	9	6
1M9U-A	70	20	8	57	43	11	5	6
1SGI-B	52	18	11	52	48	28	14	14
1A0L-A	54	24	13	63	37	21	12	9
1ABJ-H	52	18	11	52	48	28	14	14
2ANY-A	52	19	13	53	47	21	13	8
1AO5-A	54	23	10	61	39	20	11	8
1DPO-A	61	24	9	61	39	17	8	10
1BIT-A	61	20	12	58	42	15	7	8
1EQ9-A	60	25	9	53	47	23	12	11
1BQY-A	55	23	11	56	44	22	12	11
MER024901	43	24	12	45	55	31	19	12
MER017398	45	24	12	57	43	25	13	12
MER028331	53	23	11	55	45	26	13	13
MER016541	51	20	14	51	49	28	19	9
Mean ± SD	58 ± 9	21 ± 3	11 ± 2	55 ± 5	45 ± 5	21 ± 6	11 ± 4	10 ± 3
<b>Catalytic core residues</b>								
MER024901	77	23	9	45	55	23	14	9
MER017398	84	5	5	42	58	16	5	11
MER028331	68	23	18	50	50	18	14	5
MER016541	71	25	8	50	50	21	13	8
Mean ± SD	75 ± 7	19 ± 9	10 ± 6	47 ± 4	53 ± 4	20 ± 3	12 ± 4	8 ± 3

\* Amino acid physico-chemical classes: small (Gly, Ala, Ser, Pro, Val, Thr, Cys), aliphatic (Ile, Leu, Val), aromatic (Phe, His, Trp, Tyr), non-polar (Ala, Cys, Phe, Gly, Ile, Leu, Met, Pro, Val, Trp, Tyr), polar (Asp, Glu, His, Lys, Asn, Gln, Arg, Ser, Thr), charged (Asp, Glu, His, Lys, Arg), basic (His, Lys, Arg), and acidic (Asp and Glu).